```
XX
DR   WPI; 2001-316444/33.
DR   N-PSDB; AAS07029.
XX
PT   New polypeptides derived from Streptococcus agalactiae are useful to
PT   provide detection of, and vaccination against, Group B Streptococcus
PT   infections, particularly to prevent infection in neonatals.
XX
PS   Claim 1; Fig 1; 178pp; English.
XX
CC   AAU03601-AAU03722 represent Group B Streptococcus (Streptococcus
CC   agalactiae) amino acid sequences of the invention. S. agalactiae is an
CC   encapsulated bacterium which is a major pathogen of humans causing sepsis
CC   and meningitis in neonates as well as adults. The S. agalactiae antigenic
CC   polypeptides are used to vaccinate against Group B Streptococcus
CC   infections, particularly to prevent infection in new born children
CC   arising from the maternal genital tract. An immunogenic composition is
CC   useful in the preparation of a medicament for the treatment or
CC   prophylaxis of Group B Streptococcus infection. The invention does not
CC   have the disadvantages of varied response rate associated with prior art
CC   capsid polysaccharide vaccination against Group B Streptococcus
XX
SQ   Sequence 1055 AA;

  Query Match             74.4%;  Score 3777.5;  DB 4;  Length 1055;
  Best Local Similarity   71.7%;  Pred. No. 1.7e-221;
  Matches  756;  Conservative  105;  Mismatches  140;  Indels   53;  Gaps    5;

Qy          2 KKHLKTVALTLTTVSVVTHNQEVFSLVKEPILKQTQASSSISGADYAESSGKSKLKINET 61
              ||||||:|| ||||||||::|||: | :|  :|| |  |: |  |: |   : |   :::
Db          3 KKHLKTLALALTTVSVVTYSQEVYGLERESVKQEQTQSA-SEDDWFEEDNERKTNVSKE 61

Qy         62 SGPVDDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENT-ESEKQITSGSQLEQSKE 120
               :  ||:||:||||  :       |          :|  ||  | | | |:    |::|  |:
Db         62 NSTVDETVSDLFSDGNSNNSSSKTESVVSDPKQVPKAKPEVTQEASNSSNDASKVEVPKQ 121

Qy        121 SLSLNKTVPSTSNWEICDFITKGNTLVGLSKSGVEKLSQTDHLVLPSQAADGTQLIQVAS 180
                :  |      || ||  ||:|:|:||| ||||: ||||| |||||| |||||| ||||
Db        122 DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSHAADGTQLTQVAS 181

Qy        181 FAFTPDKKTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNSYLLKKVTIPTGYKHIGQ 240
              ||||||||||||||| ||||: |:|:|:| ||||:||||:|:| | |:||| ||| |||
Db        182 FAFTPDKKTAIAEYTSRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQ 241

Qy        241 DAFVDNKNIAEVNLPESLETISDYAFAHLALKQIDLPDNLKAIGELAFFDNQITGKLSLP 300
              |||||||||||||||||||||||||||||::|:|: ||||||:|||||||||||  ||| |
Db        242 DAFVDNKNIAEVNLPESLETISDYAFAHMSLKQVKLPDNLKVIGELAFFDNQIGGKLYLP 301

Qy        301 RQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEAFTGNP 360
              | |::|||||||| |:|:|| |: ||||||||||||:|   :|||||||||||||||||
Db        302 RHLIKLAERAFKSNRIQTVEFLGSKLKVIGEASFQDNNLRNVMLPDGLEKIESEAFTGNP 361

Qy        361 GDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
              ||:|||:||| |::|:|| ||||||||||||||||: :|::|||||||||||||||||
Db        362 GDEHYNNQVVLRTRTGQNPHQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421

Qy        421 GFSNKGLQKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480
              |||||||||:|||||||||||||:||||||||||||||||:||||||||||:|||||||
Db        422 GFSNKGLQKVRRNKNLEIPKQHNGITITEIGDNAFRNVDFQSKTLRKYDLEEIKLPSTIR 481

Qy        481 KIGAFAFQSNNLKSFEASDDLEEIKEGAFMNNRIETLELKDKLVTIGDAAFHINHIYAIV 540
              |||||||||||||||||:||||||||||||||||| ||:|||||: ||||||||||||||
Db        482 KIGAFAFQSNNLKSFEASEDLEEIKEGAFMNNRIGTLDLKDKLIKIGDAAFHINHIYAIV 541

Qy        541 LPESVQEIGRSAFRQNGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEQKQLTEIPVQAF 600
              |||||||||||||||||| :|:|:|:||||:|||||||::||  ::|||||| ||| ||||
Db        542 LPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFLSNKLESVNLSEQKQLKTIEVQAF 601

Qy        601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVK 660
              ||||| ||:|| :|:||||||||:||||::: :| || | ||| |   ||||::|  ||||:
Db        602 SDNALSEVVLPPNLQTIREEAFKRNHLKEVKGSSTLSQITFNAFDQNDGDKRFGKKVVVR 661

Qy        661 THHNSYALADGEHFIVDPDKLSSTIVDLEKILKLIEGLDYSTLRQTTQTQFRDMTTAGKA 720
              ||:||: ||||| ||:||||||||:||||:|||:|:||||||||||||||||:|||||||
Db        662 THNNSHMLADGERFIIDPDKLSSTMVDLEKVLKIIEGLDYSTLRQTTQTQFREMTTAGKA 721

Qy        721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNGQLLERSINKAVL 780
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        722 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNGHLLERSINKAVL 781

Qy        781 AYNNSAIKKANVKRLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY 840
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        782 AYNNSAIKKANVKRLEKELDLLTDLVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY 841

Qy        841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILN 900
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||| || |||
Db        842 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHTLAVATLADYEGLYIKDILN 901

Qy        901 SKLSQLTSIRQVPTAAYHRAGIFQAIQNAAAEAEQLLPKPGTHSEKSSSSESANSKDRGL 960
              | | :: :|||:| | ||| ||||||:|||||::||||       :  |    :|  | : : :
Db        902 SSLDKIKAIRQIPLAKYHRLGIFQAIRNAAAEADRLLPK----TPKGYLNEVPNYRKKQM 957

Qy        961 QSNPK------------------------------TN----------------RGRHSA 973
              : | |                               ||                 : :
Db        958 EKNLKPVDYKTPIFNKALPNEKVDGDRAAKGHNINAETNNSVAVTPIRSEQQLHKSQSDV 1017

Qy        974 ILPRTGSKGSFVYGILGYTSVALLSLITAIKKKK 1007
              ||:| || :|:| |||| |: || |:|| || |
Db       1018 NLPQTSSKNNFIYEILGYVSLCLLFLVTAGKKGK 1051


RESULT 5
ABP56257
ID   ABP56257 standard; protein; 1055 AA.
XX
AC   ABP56257;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Serotype III group B Streptococcus strain COH1 BVH-A4 SEQ ID NO:2.
XX
KW   Serotype III group B streptococcus strain COH1; BVH-A4; streptococcus;
KW   antibiotic; immunostimulant; vaccine; bacterial infection; sepsis;
KW   meningitis; pneumonia; cellulitis; osteomyelitis; septic arthritis;
KW   endocarditis; epiglottitis; osteomyelitis; amniotis; endometritis;
KW   cellulitis; fasciitis; bacteraemia; urosepsis; peritonitis; emphysema;
KW   mastitis; streptococcal infection.
XX
OS   Streptococcus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .22
FT                   /label= signal
FT   Protein         23. .1055
FT                   /label= BVH-A4
XX
PN   WO200288178-A2.
XX
PD   07-NOV-2002.
XX
PF   02-MAY-2002; 2002WO-CA000664.
XX
PR   02-MAY-2001; 2001US-0287712P.
XX
PA   (SHIR-) SHIRE BIOCHEM INC.
XX
PI   Martin D,  Hamel J,  Brodeur BR,  Rioux S,  Boyer M;
XX
DR   WPI; 2003-120461/11.
DR   N-PSDB; ABZ21973.
XX
PT   New BVH-A4 proteins and genes from serotype III Group B streptococcus,
PT   useful for treating or preventing streptococcal infection in infants,
PT   pregnant women, non-pregnant adults (e.g. pneumonia), or members of dairy
PT   herd (mastitis).
XX
```

197	3.9	2441	2	O96124	O96124 plasmodium
197	3.9	4007	2	Q7SHZ4	Q7shz4 neurospora
96.5	3.9	1680	2	Q9U8G1	Q9u8g1 plasmodium
96.5	3.9	1875	1	MLP1_YEAST	Q02455 saccharomyc
96.5	3.9	2139	2	Q07569	Q07569 entamoeba h
96.5	3.9	2479	2	Q66GS8	Q66gs8 homo sapien

ALIGNMENTS

```
20      PRELIMINARY;      PRT;  1008 AA.
20;
JN-2001 (TrEMBLrel. 17, Created)
UN-2001 (TrEMBLrel. 17, Last sequence update)
AR-2004 (TrEMBLrel. 26, Last annotation update)
thetical protein SPy0843.
redLocusNames=SPy0843;
ptococcus pyogenes.
eria; Firmicutes; Lactobacillales; Streptococcaceae;
ptococcus.
_TaxID=1314;

JENCE FROM N.A.
AIN=SF370 / ATCC 700294 / Serotype M1;
LINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
retti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
neaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
n Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J.,
n X., Clifton S.W., Roe B.A., McLaughlin R.E.;
mplete genome sequence of an M1 strain of Streptococcus pyogenes.";
c. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
L; AE006534; AAK33772.1; -.
 GO:0009986; C:cell surface; IEA.
erPro; IPR001899; Gram_pos_anchor.
erPro; IPR007093; LRR_Tp.
SITE; PS50847; GRAM_POS_ANCHORING; 1.
plete proteome; Hypothetical protein.
UENCE   1008 AA;  111503 MW;  045BF8CE931AF0CF CRC64;

Match              100.0%;  Score 5080;  DB 2;  Length 1008;
Local Similarity   100.0%;  Pred. No. 2.4e-215;
es 1008;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

  1 MKKHLKTVALTLTTVSVVTHNQEVFSLVKEPILKQTQASSSISGADYAESSGKSKLKINE 60
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MKKHLKTVALTLTTVSVVTHNQEVFSLVKEPILKQTQASSSISGADYAESSGKSKLKINE 60

 61 TSGPVDDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENTESEKQITSGSQLEQSKE 120
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 61 TSGPVDDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENTESEKQITSGSQLEQSKE 120

121 SLSLNKTVPSTSNWEICDFITKGNTLVGLSKSGVEKLSQTDHLVLPSQAADGTQLIQVAS 180
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 SLSLNKTVPSTSNWEICDFITKGNTLVGLSKSGVEKLSQTDHLVLPSQAADGTQLIQVAS 180

181 FAFTPDKKTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNSYLLKKVTIPTGYKHIGQ 240
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 FAFTPDKKTAIAEYTSRAGENGEISQLDVDGKEIINEGEVFNSYLLKKVTIPTGYKHIGQ 240

241 DAFVDNKNIAEVNLPESLETISDYAFAHLALKQIDLPDNLKAIGELAFFDNQITGKLSLP 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 DAFVDNKNIAEVNLPESLETISDYAFAHLALKQIDLPDNLKAIGELAFFDNQITGKLSLP 300

301 RQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEAFTGNP 360
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 RQLMRLAERAFKSNHIKTIEFRGNSLKVIGEASFQDNDLSQLMLPDGLEKIESEAFTGNP 360

361 QDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 QDDHYNNRVVLWTKSGKNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

Qy      421 GFSNKGLQKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 GFSNKGLQKVKRNKNLEIPKQHNGVTITEIGDNAFRNVDFQNKTLRKYDLEEVKLPSTIR 480

Qy      481 KIGAFAFQSNNLKSFEASDDLEEIKEGAFMNNRIETLELKDKLVTIGDAAFPHINHIYAIV 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 KIGAFAFQSNNLKSFEASDDLEEIKEGAFMNNRIETLELKDKLVTIGDAAFPHINHIYAIV 540

Qy      541 LPESVQEIGRSAFRQNGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEQKQLTEIPVQAF 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 LPESVQEIGRSAFRQNGANNLIFMGSKVKTLGEMAFLSNRLEHLDLSEQKQLTEIPVQAF 600

Qy      601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVK 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 SDNALKEVLLPASLKTIREEAFKKNHLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVK 660

Qy      661 THHNSYALADGEHFIVDPDKLSSTIVDLEKILKLIEGLDYSTLRQTTQTQFRDMTTAGKA 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 THHNSYALADGEHFIVDPDKLSSTIVDLEKILKLIEGLDYSTLRQTTQTQFRDMTTAGKA 720

Qy      721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNGQLLERSINKAVL 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 LLSKSNLRQGEKQKFLQEAQFFLGRVDLDKAIAKAEKALVTKKATKNGQLLERSINKAVL 780

Qy      781 AYNNSAIKKANVKRLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AYNNSAIKKANVKRLEKELDLLTGLVEGKGPLAQATMVQGVYLLKTPLPLPEYYIGLNVY 840

Qy      841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILN 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 FDKSGKLIYALDMSDTIGEGQKDAYGNPILNVDEDNEGYHALAVATLADYEGLDIKTILN 900

Qy      901 SKLSQLTSIRQVPTAAYHRAGIFQAIQNAAAEAEQLLPKPGTHSEKSSSSESANSKDRGL 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 SKLSQLTSIRQVPTAAYHRAGIFQAIQNAAAEAEQLLPKPGTHSEKSSSSESANSKDRGL 960

Qy      961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 QSNPKTNRGRHSAILPRTGSKGSFVYGILGYTSVALLSLITAIKKKKY 1008
```

```
RESULT 2
Q8P1F7
ID   Q8P1F7      PRELIMINARY;      PRT;  1008 AA.
AC   Q8P1F7;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein spyM18_0903.
GN   OrderedLocusNames=spyM18_0903;
OS   Streptococcus pyogenes (serotype M18).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=186103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MGAS8232;
RX   MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA   Smoot J.C., Barbian K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA   Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA   Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA   Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT   "Genome sequence and comparative microarray analysis of serotype M18
RT   group A Streptococcus strains associated with acute rheumatic fever
RT   outbreaks.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR   EMBL; AE010019; AAL97553.1; -.
DR   GO; GO:0009986; C:cell surface; IEA.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR007093; LRR_Tp.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Complete proteome; Hypothetical protein.
```